DOCUMENT RESUME

ED 340 766 TM 018 022

AUTHOR	Venezky, Richard L.
TITLE	Assessing Higher Order Thinking and Communication Skills: Literacy.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
PUB DATE	30 Sep 91
NOTE	44p.; Commissioned paper prepared for a workshop on Assessing Higher Order Thinking & Communication Skills in College Graduates (Washington, DC, November 17-19, 1991), in support of National Education Goal V, Objective 5. For other workshop papers, see TM 018 009-024.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Adults; Cognitive Measurement; Communication Skills; *Critical Thinking; *Educational Assessment; *Evaluation Methods; Evaluation Problems; Higher Education; *Literacy; Reading Comprehension; *Thinking Skills; Writing Evaluation
IDENTIFIERS	*Higher Order Learning; *National Education Goals 1990

ABSTRACT

Assessment of higher order literacy skills encounters three initial problems aside from assessment methods: (1) definition of literacy; (2) range of skills to assess; and (3) whether or not higher order literacy can be assessed independently of a particular content area. Regardless of definitions, the general performance areas to be covered must be decided. Issues to be resolved are the question of how many lower-level literacy skills to include in a higher-level assessment and the degree to which a pure literacy assessment can be separated from any particular content. The cognitive base for assessment of adult literacy is thin, and that dealing with higher-level skills is almost non-existent. In the future, text-based tasks might be developed from analyses of basic reading comprehension processes. These tasks might be aligned with the Delphi classification of core critical thinking skills. To be compatible with the basic directions of critical thinking and communications research, it is suggested that literacy assessment be built around a definition of human expertise, confined primarily to reading and writing, and integrated with content-area assessments, at least for areas where text-based information must be integrated with information previously acquired. Three figures illustrate the discussion, and a 52-item list of references is included. Reviews by R. Calfee, M. A. Miller, and M. Scriven of this position paper are provided. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340766

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Assessing Higher Order Thinking and Communication Skills: Literacy

Richard L. Venezky
University of Delaware
September 30, 1991

TM018022

Position paper prepared for the National Center on Educational Statistics Workshop on Assessing Higher Order Thinking and Communication Skills in College Graduates, November, 1991. Do not quote without permission.

BEST COPY AVAILABLE

Abstract

Introduction to problems

Assessment of higher order literacy skills encounters three initial problems, aside from assessment methods: (1) the definition of literacy, (2) the range of skills to assess, and (3) whether or not higher order literacy can be assessed independently of a particular content area. Most current definitions of literacy arose out of a concern for basic literacy ability and derive from a functional anlysis of social demands. In contrast, current analyses of critical thinking are based on human expertise--the cognitive skills found to underlie competent performance.

However this issue is resolved, the general performance areas to be covered must be decided. Inflation has now invaded the functional literacy area, with reading, writing and basic arithmetic already part of the standard definition and communication and interpersonal skills waiting for entry. Beyond these issues is the question of how many of the lower-level literacy skills to include in a higher-level assessment. Finally, the degree to which a "pure" literacy component can be separated from any particular content needs to be resolved.

Summary of Current Knowledge

The cognitive base for assessment of adult literacy is thin; that which deals with higher-level literacy skills almost non-existent. Some statistical work on information location has been done using data from recent adult literacy surveys and an older literature exists on selected topics, such as reading - listening differences, reading speed vs. content area, and proxy measures.

Suggestions for Development and Implementation

Text-based tasks might be developed from analyses of basic reading comprehension processes, i.e., identifying, analyzing, and synthesizing information. With some squeezing, these tasks might be aligned with the Delphi classification of core critical thinking skills. Higher levels of text analyses that require integration of text-derived knowledge with previously acquired

knowledge can also be developed for assessment.

Arguments for/against

To be compatible with the basic directions of critical thinking and communications research, it is suggested that literacy assessment be built around a definition of human expertise, that it be confined primarily to reading and writing, and that it be integrated with content-area assessments, at least for the levels wherein text-based information must be integrated with information previously acquired.

Assessing Higher Order Thinking and Communications Skills: Literacy

Introduction to Problems

Definition of literacy

The first issue in designing a literacy assessment for higher level thinking and communications skills centers on the definition of literacy itself. Since World War II most of the operative definitions of literacy have been functional, beginning with the UNESCO reports of the early and mid 1950s. (UNESCO, 1957; Gray, 1956) UNESCO's concern with literacy derived from its efforts in fundamental education, which was defined as "that kind of minimum and general education which aims to help children and adults who do not have the advantage of formal education to understand the problems of their immediate environment..." (Gray, 1956, pp. 16f). From this focus on minimal levels of education, a somewhat tautological definition of functional literacy was developed.

> "...a person is functionally literate when he has acquired the knowledge and skills in reading and writing which enable him to engage effectively in all those activities in which literacy is normally assumed in his culture or group" (ibid.).

Functional definitions since the UNESCO publications have continued to emphasize both competence with printed information and socially appropriate deployment of this ability. For example, the National Assessment of Educational Progress, for its survey of young adult literacy skills, defined literacy as "using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential" (Kirsch & Jungeblut, 1986, p. I-8). This is also the definition of literacy adopted for the current National Adult Literacy survey, sponsored by NCES. Similarly, Statistics Canada recently defined literacy for its national survey

of adult literacy as "The information processing skills necessary to use the printed material commonly encountered at work, at home, and in the community" (Neice & Adsett, 1991, p. 3). In a survey and discussion of literacy definitions, Venezky (1990) suggested adoption of two levels of literacy: basic literacy, which is that level of literacy needed to be self sustaining in acquiring further literacy, and required literacy, which is that level of literacy required by one's societal aspirations.

From the functional view of literacy, a testing tradition has developed wherein items are selected from a matrix formed by crossing the processing demands of different areas of social functioning (e.g., work, citizenship, home management) with either life skills or document types (e.g., prose texts, line graphs, signs). Within the resulting cells, items are selected to represent tasks and difficulty levels typical of each intersection. This procedure characterizes earlier literacy surveys done by Louis Harris and Associates (1970, 1971), the Adult Performance Level Project (Northcutt, 1975), and the National Assessment of Education Progress (1976), as well as the Young Adult Literacy Survey (Kirsch & Jungeblut, 1986) and the National Adult Literacy Survey now underway.

An alternative to a functional definition of literacy is a cognitive definition that attempts to define literacy in terms of levels of expertise for human performance. The groundwork for approaching literacy as a cognitive process was laid early in this century by a psychologist whose name is rarely invoked in discussions of cognitive processing, namely Edward Thorndike. From a series of studies on paragraph reading errors, Thorndike concluded that reading involved a high degree of reasoning. "It thus appears that reading an explanatory or argumentative paragraph in his textbooks on geography or history or civics, and (though to a less degree) reading a narrative or description, involves the same sort of organization and analytic action of ideas as occur in thinking of supposedly higher sorts" (Thorndike, 1917, pp. 331). A step toward defining competencies for assessment was taken by Neijs (1961), who defined the mature attainment of literacy in terms of basic reading abilities: a reading vocabulary equal to or greater than the 2,500 words most

frequently encountered in print, independent word recognition skills, a silent reading speed equal to or greater than 150 words per minute, and reasonably effective oral reading. Curiously, comprehension ability is not overtly mentioned in this definition.

Most of the current work, as limited as it is, on adult literacy as a cognitive process is compatible with the criterion-referenced measurement movement that originated with the work of Robert Glaser and his colleagues in the early 1960s (e.g., Glaser & Klaus, 1962). Although adult literacy research is still focused on the lower levels of functioning and still subscribes mostly to functional definitions of literacy, more cognitively based definitions can be derived from recent work on reading and writing (e.g., Pearson, 1986; Baker, Freeman & Clayton, 1991; Glaser, 1991).

Skill types

The second issue in higher-order literacy assessment centers on the types of skills to assess in college-level literacy. Several hundred years ago reading ability alone determined literacy. Since at least the middle of the nineteenth century, however, both reading and writing have been considered necessities for a literate individual. More recently, dating mostly since from World War II, literacy has included not only reading and writing, but basic levels of arithmetic ability needed to interpret common texts. Even within these three areas, however, issues of assessment coverage exist. Should a college-level literacy assessment include vocabulary assessment? Should it include separate assessments of charts, graphs, tables, and the like? And should it include creative writing as well as expository writing?

Of more general interest is the degree to which a test of higher level thinking and communication skills should assess lower level skills. The Young Adult Literacy Survey (Kirsch & Jungeblut, 1986) showed that some college graduates lacked basic reading and writing skills. Should the current enterprise be concerned with skills typically acquired before secondary school?

Separate vs embedded assessment

A third problem centers on both the necessity and practicality of assessing literacy as an

independent set of higher order thinking and communicating skills. The alternative is to imbed literacy assessment within separate assessments for different curricular areas or within assessments for specific higher-order skills. The central issue here is the degree to which higher order literacy skills can be assessed without dependence on specific background knowledge. Some current work in curricular area assessment is nearly identical to work in reading comprehension. For example, Baker, Freeman & Clayton (1991) approaches the assessment of secondary level history as an assessment of reading comprehension where primary emphasis is placed on the integration of knowledge read from a test passage with previously acquired knowledge. Student essays are evaluated according to two independently developed scales: one for writing ability and one for content knowledge. Within the curricular area approach, however, there may be a logical trap. Imagine, for example, text based assessments in literature or geology or history, wherein students read passages in these content areas and then produced responses based at least in part on what they read. (Whether these are long or short passages, whether the responses are open-ended or constrained by multiple choices, or any of the issues related to item construction make no difference here.) Assume, furthermore, that two scoring scales are developed: one for the content area and one for a set of "pure" literacy skills.

Although writing ability might be assessed thusly, no direct assessment of reading is possible under such a plan unless questions are included which tap reading comprehension independently of content area knowledge and this is equivalent to generating a separate literacy assessment. An escape from this predicament exists, but it requires further research. Information input for each student could be through both reading and listening, where items would be drafted in pairs with equivalent passages and production tasks (each item would require only one input process: either reading or listening.) Reading ability would then be scored according to the difference in performance levels between aural and print input.

Summary of Current Knowledge

The knowledge base for testing higher levels of literacy is composed of a disparate collection of work, located more on the periphery than in the center of the area of concern. Studies of reading comprehension over the past 100 years have focused mostly on the k-8 school range, with a small number reaching into the secondary levels and even a smaller number examining reading at the post-secondary levels. Three subtopics will be discussed here: assessment of adult literacy, adult reading processes, and proxy measures for adult literacy.

Assessment of adult literacy

Current approaches to surveying adult literacy abilities derive from the Adult Performance Level Study done at the University of Texas in the 1970s (Final Report, 1977). In the summer of 1971 the Texas State Education Agency received a contract from the Division of Adult Education of the U. S. Office of Education to develop a definition of adult functional literacy as a basis for an adult education system. The new definition of adult literacy was to be developed through " an analysis of adult living rather than the common practice of attaching a grade equivalence to [the requirements of adult living]." (Final Report, 1977; cited in Fischer, Haney, & David, 1980, p. 57). The actual research that led to the Adult Performance Level test was carried out by the Division of Extension of the University of Texas at Austin, under subcontract to the Texas State Education Agency.

From a matrix of basic skills (e.g., communication, computation, problem solving) and general knowledge areas (health, government and law, transportation), test items were developed and administered to a sample of approximately 4,000 adults, selected primarily from adult education programs across the nation. In the second phase of the study, the framework and test items were revised and a second field testing carried out with about 7,500 adults. Three competency levels were defined for reporting of the results, designated simply as APL levels 1, 2, and 3. Following conclusion of the second phase of the study, the American College Testing (ACT) Program received the rights to refine and publish the APL tests. The APL Program, as published by ACT, consists of an Adult Performance Level General Survey (40 multiple-choice items) and five Content Area Measures (community resources, occupational knowledge, consumer economics, health, and government and law).

Criticisms of the APL project have centered on its approach to literacy definition, its testing methodology, and its reporting of results. Griffith and Cervero (1977), for example, criticized the emphasis on survival skills rather than success skills, and the arbitrary nature of the three APL levels and their use in defining "functional incompetence." Nafziger et al. (1975) was particularly critical of the technical aspects of the APL tests, finding extremely low levels of concurrent and

predictive validity and equally low levels of test reliability. Further technical criticisms are summarized by Fischer, Haney, and David (1980).

Although other attempts to define and assess adult literacy skills have been made since the APL study , the most important effort for the present project are the various NAEP adult literacy assessments that began with the Young Adult Literacy Survey of 1985 (Kirsch & Jungeblut, 1986). Under contract to ETS, NAEP has developed a far more sophisticated approach to testing than was used in the APL study. Using BIB spiraling to allow a larger item pool than what any single subject encounters, and IRT scaling to relate item difficulties to subject competencies, a series of literacy tests have been developed with sufficient overlap of items to allow outcome linkage.

The item development matrix for the Young Adult Study was based upon text types (12 categories) and processing demands. Within the matrix, items with different levels of difficulty were created by varying the task requirements. Included within the task and the required response were various combinations of reading, writing, listening, and speaking. In the original assessment design, processing demands were classed as knowledge, evaluation, specific information, social interaction, and application. The processing demands of the exercises included in the final test design can be organized in four classes: location or entry of facts, analysis of groups of facts, interpretation, and summarizing (Venezky, Kaestle, & Sum, 1987). In factor analyses done on the results from approximately 4,000 persons, ages 21-26, three literacy performance scales were defined: prose, document, and quantitative.

Since the release of the Young Adult Literacy Survey results, various studies have been done to explore the cognitive processes underlying the results obtained (e.g., Kirsch & Mosenthal, 1988, 1990; Guthrie, 1988; Sheehan & Mislevy, 1990). Kirsch and Mosenthal (1990), for example, developed a parsing scheme for 61 of the document scale tasks along with their related documents. Document, task, and processing variables were then identified that could account for 89% of the total variance in the individual responses. Document and task variables included such factors as number of organizing principles and number of specifics; processing variables included (1) degrees of correspondence between task specification and document statement; (2) types of information required by the task, and (3) plausibility of distractors in the text. Considerable work remains, however, in accounting for the results obtained across all three scales.

A newer trend in elementary and secondary level assessment, called process assessment, is reflected in several state-wide assessment programs (e.g., Illinois, Maryland, Michigan). In the State of Maryland, for example, reading assessment consists of a variety of extended tasks, based upon two long texts, one narrative and one expository. Testing is done in a number of lengthy sessions, where students must reflect upon the texts and do various summarizing, analysis, and

creative exercises. Since results from the various state-wide assessments are only now being published, little empirical work is available on this approach. It does appear to be important for the assessment of higher levels of literacy and may provide a model for portions of a college-level assessment.

Adult reading processes

Reviews of research on adult literacy have tended over the past three decades to point out as many problems with research in this field as they do usable results (e.g., Brunner et al., 1959; Weber, 1975; Venezky, 1991). While a considerable literature exists on the comprehension skills of school-age children (see Pearson, 1986; Barr, et al., 1990 for reviews), much less is know about the processing habits of adults. Models of reading development exist for children (e.g., Gray, 1937; Chall, 1983), but not for adults as new learners and not for the critical levels of reading ability that extend beyond what the high school student might be able to do.

One topic for which an interesting knowledge base exists is the comparison of listening and reading ability. Although, as described above, suggestions have been made to use the difference between listening comprehension and reading comprehension as an indicator of teachability, the relationship between these two abilities is far from simple. Goldstein (1940), for example, found that with adults (ages 18 to 65):

1. Superiority of listening over reading increases with decreasing difficulty of text. That is, the easier the text is to understand, the greater will be the advantage for listening over reading;
2. As intelligence decreases, the advantage of listening over reading increases.; and
3. Passages that test equivalently in reading tasks may not do so in listening tasks.

In other words, the use of a listening as an indicator of an upper bound for reading ability would need to be done in relationship to both text difficulty and to intelligence. The possible use of this differential, however, to isolate a "literacy" factor for higher level skills may be possible, but as indicated above, more work is required to determine if this is justified.

The role of background knowledge in adult reading comprehension and reading speed has been an interest to reading researchers for at least 70 years (e.g., Judd & Buswell, 1922; Dixon, 1951; Birkmier, 1982). In general, the more careful studies find that among college graduates, there are no intrinsic subject matter differences. While subjects will read materials from their own areas faster and more accurately than materials from other areas, no single area will consistently show superior results when subjects are randomly selected from the areas and matched for IQ.

Of more recent interest is the difference between situated and decontextualized testing and learning. The current interest in situated learning (e.g., Rogoff, & Lave, 1984; Collins, Brown & Newman, 1989) has struck a favorable chord for adult literacy educators who have found, in

general, that adult learners respond best to instruction that is situated within familiar life contexts. However, Scribner and Cole (1981) found that among the Vai people of Liberia, literacy acquired within a school setting (and therefore derived from generally decontextualized instruction) resulted in far superior cognitive consequences to literacy acquired in natural settings. Exactly what makes learning difficult for many adults, that is, the abstractness of most school-based tasks, may be what is critical for acquiring higher level thinking skills. Goody and Watt (1968) made a related argument in surveying the history of writing and literacy. In their conceptual framework for the evolution of literacy, writing has cognitive consequences exactly because it allows reflection on decontextualized knowledge.

This is a far more complex issue than can be discussed here; it is, nevertheless, an important one for the design of college-level literacy assessment. If critical thinking skills are to be assessed and if abstract thinking ability is to be counted among these skills, some part of the assessment scheme then must include tasks in which knowledge is decontextualized. That is, while some part of a literacy assessment might be built upon ordinary, every-day tasks, assessment of critical ability will require a departure from this technique.

A final issue to note, but not to elaborate on, is reading strategies. Exploration of literacy as a human competence depends strongly on the analysis of reading strategies of expert readers (and writers). An interesting literature has developed over the past decade on various aspects of reading strategies, particularly in children (e.g., Brown, 1987; Paris, Lipson, & Wixson, 1983). The examination of reading strategies of adults has been more limited. Among the few studies in this area is a survey by Mikulecky(1982) of the reading tasks and reading strategies of high school juniors, adult technical school students, and adult workers from a variety of occupations. The workers, for example, differed from the others in their more frequent use of underlining and note taking. Older work has examined the ability of adults to vary their reading strategies according to task and content demands (e.g., Bond & Bond, 1941). Most of these studies, however, did not isolate cognitive strategies; instead they focused on behavioral factors such as on reading speed, note taking, and the like.

Proxy measures

Proxy measures for adult literacy were used as early as 1840 by the U. S. Census , which asked individuals if they could read and write (Stedman & Kaestle, 1986). Although the specific questions asked by the Census have varied over time, proxy measures for literacy have remained in

the Census to the present day. UNESCO, as part of its functional literacy assessment, adopted years of education as a proxy for literacy (Gray, 1956; UNESCO, 1957). Countries have continued to supply literacy data to UNESCO in terms of this measure. For example, the Canadian Commission for UNESCO adopted in 1983 the criterion of less than five years of formal schooling as an indicator of *basic illiteracy* and less than nine years as an indicator of *functional illiteracy* (Thomas, 1983). The most extensive study of proxy measures for adult literacy, using empirically derived data, is Neice and Adsett (1991), which is based upon a recent Canadian adult literacy survey.

The Canadian survey, carried out by Statistics Canada in the late 1980s, was based upon the NAEP Young Adult Literacy Survey. In the proxy study, results from 9,455 persons sampled across the Canadian provinces, ages 16-69, were used to test various proxy measures which had been included in the background survey given to each individual who was tested. Educational attainment, particularly as defined by the Canada/UNESCO criterion, was not a particularly accurate predictor; almost 44% of those who would have been classed as "basic illiterates" by this criterion tested at a higher level of ability. Self-assessment proved to be more powerful, while frequency of reading books was (potentially) the single best indicator.

Magazine reading was also a fairly good predictor, but tv and radio listening/watching were not because too many people did them. A factor analysis revealed two principal components for literacy prediction: a *behavioral component*, which included educational attainment, frequency of library visits, and frequency of letter writing, and an *assessment component*, which included the direct and self-assessments (e.g., "Do you feel your reading skills in English (French) are adequate for this job?") The authors concluded that (1) high school completion is the minimum necessary education for adequate reading skills in North America, and (2) self-assessment indicators are not adequate by themselves; they must be combined with behavioral indicators (e.g., frequency of library visits).

Keep in mind that the proxy measures used in Neice and Adsett (1990) were developed for a broad scale of literacy ability; they were not oriented toward higher or critical literacy and probably would not discriminate adequately at that end of the scale. Nevertheless, the study may provide a model for the development of proxies for critical literacy, particularly through its use of statistical modeling.

Suggestions for Development and Implementation

In this section I will consider how an assessment of literacy could be built from a model of expert behavior, considering two general classes of skills: text-based and integration of text derived knowledge with previously acquired knowledge. For the former I will first present some ideas about assessment of basic reading comprehension skills, adapted from an unpublished background report developed for the 1979 NAEP reading and literature assessment. Then, assessment items are considered, based upon one of the critical thinking skill categories developed for the American Philosophical Association's Committee on Pre-College Philosophy (Facione, 1990). For integration of text-derived information with previously acquired information, I will draw upon political and historical analyses of a familiar American text. In choosing these categories of assessment to explore, I am clearly choosing in favor of a literacy assessment based upon analyses of human expertise rather than upon socially-relevant literacy tasks. The reasons for this decision are given in the concluding section of this paper.

Text-based tasks

Text-based literacy assessment assumes that a minimal amount of previously acquired knowledge is required for a specified task. By minimal is meant something different from units or amounts of information. A more precise definition might be that the tasks should require no information beyond what is required for lexical and syntactic analysis and recognition of the text propositions. Even this description, however, is imprecise in that it does not clarify where common, everyday knowledge or schema separate from specialized knowledge in the making of inferences, the resolution of ambiguity, and the like. For the present I will not attempt to resolve this issue any further, other than through the examples offered below.

In all cases, difficulty level of a task can be manipulated by various elements within the task statement, the text, and the relationship between the text and the task statement. Kirsch and Mosenthal (1990) have identified such elements in their analysis of information location tasks from the Young Adult Literacy Survey

Basic comprehension skills

Reading tasks can be classed according to the operations which each requires. For the present discussion the following tasks are defined.

1. Identifying: These operations produce answers to such questions as "When was Abe Lincoln

born?", "Where did the cat land when it fell?" and "Did Nimitz score the winning goal?" Identifying involves matching of components in a question to similar or identical components in a text, either for verifying premises given in the question or finding information associated with the components sought for in the text. Typically, only a small amount of text is required for answering Identifying questions.

The difficulty of these operations depends upon:

a. The amount of recoding required to determine what the question is seeking.

b. The similarity between the key question components and the related text components.

c. The number and types of distractors in the text.

d. The relationship between additional information which is sought and the key text components to which this information relates.

e. The complexity of the question in relation to the text.

To illustrate how difficulty of identifying operations might vary, consider the following task.

Question 1: Did Jimmy eat the sandwich because he was hungry?

Text 1. Jimmy ate the sandwich because he was hungry.

In this item, verification is required. The key components in the question (Jimmy (actor), ate (action), sandwich (object of action), hungry (cause for action) match almost exactly the key components in the text. No additional information is sought. In general, this would be below the difficulty level desired for a college exit examination.

A more difficult form of this task might have the same question, but either of the following texts.

Text 2. Jimmy saw a sandwich on the shelf. He hadn't touched food for almost 18 hours. He looked around once, reached onto the shelf, and gobbled down the treat.

Text 3. That Jimmy ate the sandwich was certain. Why he did it was a different issue. Some say because he never refuses food. Others believe it was strictly a matter of unrequited appetite. I'm not sure myself.

Text 2 requires a minor inference, i.e., that not touching food for 18 hours makes one hungry. In addition, the key components in the text must be linked to the key components in the question through a variety of synonym and set relations (e.g., he - Jimmy, eat - gobbled, sandwich - treat).

Text 3 contains, besides linkage problems, a more difficult question-text interaction; i.e., the text does not give a simple yes-no answer to the question.

We might also achieve a high level of task difficulty by retaining Text 1, but substituting either of these questions:

Question 2: What was the motive for the male protagonist's action concerning the edible substance mentioned in the following paragraph.

Question 3: With what certainty does the following paragraph indicate hunger as the cause of Jimmy's eating the sandwich.

Both questions 2 and 3 are more difficult to encode than question 1 and require more complex operations for linking their key components to the key text components. If, however, question 2 can be encoded properly, and if hunger can be recognized as a legitimate cause, then question 2 can be answered with what is given in Text 1. But question 3 requires special reader-supplied information; viz., a judgment of degree of certainty.

2. Analyzing: These operations are used to order objects or events, select probable outcomes, verify arguments, and determine meanings based upon context. Some typical questions that might require analysis are listed below (without the text which would follow each).

a. List the steps required for building a birdhouse, in the order in which they would be done.

b. Find how Anna and Nicolina are related to each other.

c. Which of the following is the most probable:

(1) Caligula will abdicate the emperorship to Claudius so that Claudius can be a participant-observer in the office of Roman emperor.

(2) Messalina will invite Caligula to see her Thracian etchings.

(3) Claudius will have his History of Carthage re-issued, with drawings of grapes rather than elephants.

d. Based strictly upon the evidence given in the paragraph below, is the earth flat?

e. Determine from the information given which floor in Neiman-Marcus (Dallas) one would find a thyrsus.

These operations require the reader to analyze a large section of text to produce an outcome drawn either from the question or the text itself. They differ from identifying operations in that (1) their domain is potentially the entire text and (2) they require cognitive operations which generally are more complex than identifying operations.

The difficulty (or complexity) of an analysis item will vary with the same parameters identified for identifying, but in addition will depend upon the complexity of the analysis operations required.

3. Synthesizing: These operations require analysis of a text, plus generation of a response that differs from both the question and the text. Included in this category are paraphrasing and outlining.

Identifying Operations

A. Verification

Task: Determine whether or not a proposition given in a question occurs in the text. The simplest form of identifying requires a match between a question proposition and a text proposition In the example shown below, Text 1 allows such a direct match. If the key components (Sissy, whooping cranes, fly, ridge) are properly matched, the question can be answered.

Text 2 presents a slightly more difficult text encoding task, in that a pronoun (them) must be linked to its antecedent (whooping cranes). In addition, more propositions must be encoded in this sample than in text 1 before the match between question components and text components can be made.

Text 3 requires linking synonyms (sailing-fly) and separating relevant from irrelevant material before the question components and text components can be matched. Then a deduction must be made; viz., that if Sissy fell asleep just before the cranes flew over the ridge, then she didn't see them fly over.

Text 4 requires operations similar to those required for text 3, but more text must be encoded before the relevant proposition can be extracted.

Example

Question: Did Sissy see the whooping cranes fly over the ridge?

Text 1: Sissy did not see the whooping cranes fly over the ridge.

Text 2: When the whooping cranes came flying over the ridge, Sissy was sound asleep and therefore didn't see them.

Text 3: Sissy and Delores fell sound asleep just before the giant whooping cranes came sailing over the ridge.

Text 4: Sissy had hitchhiked all day in order to reach the ranch by sundown. She and Delores renewed their friendship well into the night, but by dawn both were in a deep slumber. Neither heard the ranch bell ring to announce that the whooping cranes were flying in over the ridge

B. Locating New Information

Task: Find simple information in a text by linking key components of a question to key components in a text, and then by linking the text components to the desired information. The difficulty of matching key components may be varied as described above for verification. Difficulty of linking new information will depend upon the type of linkage involved and the amount (and type) of world knowledge which the reader must supply. For assessment purposes, linkage types might include:

a. set relations
b. pronominal reference
c. cause and effect
d. synonym
e. enablement

Example A

Question: Why did Cinderella leave the party at midnight?

Text 1: Cinderella left the party at midnight because she promised her babysitter she would be home by 12:30 a.m.

Text 2:

Cinderella had arranged a clandestine meeting with Prince Chanly at the party, but to her surprise both Narding and Jasper came as guests of her ex-husband, Count Stanig. By the time the clock struck twelve she had lost all hope of realizing her tete-a-tete. Leaving one of her imitation Tiffany slippers behind as a temptation to Chanly, she slipped unnoticed into the enshrouding mist.

To do this task, the reader must first decide what information is sought and what the relevant (or key) components are for locating this information. This particular question calls for a motive for a particular action (leaving the party), performed by an actor (Cinderella) at a particular time (midnight). The action-actor-time values form a search key for locating relevant text premises.

In text 1 a simple, direct match can be made between question and text components. Linking the text components to the desired information then follows, assisted by the text connector "because." In text 2, on the other hand, the matches for the key components of the question are distributed across a group of sentences and must be matched through synonym and pronoun relationships; then, the motive for her departure must be inferred.

Example B

Question: Who was the first president of the United States?

a. John Dean
b. Eliahu Hakim
c. George Washington
d. Thomas Jefferson

Text 1: As everyone knows, George Washington was the first president of the United States.

Text 2: As everyone knows, George Washington was the first elected president of the United States. However, technically speaking, Eliahu Hakim, first Speaker of the House of Representatives, was the first president. He served, as required by the Constitution, from the time he was elected Speaker until the time that the vote for president became official--a total period of 18 minutes and 42 seconds!

Text 1 represents an exceedingly simple match for the question components and a simple statement of the desired information. Text 2 illustrates a special type of item difficulty, created by the inclusion of critical information which is contrary to common knowledge. The careless reader might select George Washington either without reading the passage, or after a cursory scan. This is the negative use of reader-supplied knowledge.

Analyzing Operations

A. Ordering

Task: Order two or more objects or events, using the appropriate dimensional system. The most common ordering tasks involve events related in time, objects related in one- or two-dimensional space, or people related by kinship linkages.

Example

Question: What was Gushkin's relationship to Andropov?

Text: Gushkin's mother, Andranova, was Alexis's half-sister by a second marriage to Nero. Andropov's great-aunt, Tantamrova, was Nerovovich's second-cousin by a tenuous link through his uncle, Hamoratnoff.

This typical white Russian family description represents how text complexity can affect an ordering task. In general, the difficulty of an ordering task will depend upon the logical complexity of the text-described relationships. This difficulty might be quantified by the number of units intervening between the two related objects.

For ordering two elements, we have aRb where a and b are elements and R is an ordering relationship. An ordering task requires finding one of these three elements given the other two.

For example, given Charlotte and Nellie, related as mother and daughter, respectively, we could have any of the following:

1. Who was Charlotte's daughter?
2. Who was Nellie's mother?
3. How was Nellie related to Charlotte?

For three or more elements to order, in general we would ask only for the appropriate ordering.

B. Predicting outcomes

Task: Given several possible outcomes for a passage, select the one that is most consistent with the premises given in the passage.

The parameters for predicting difficulty of outcome tasks have not been worked out yet. In general, predicting outcomes, when assessed in a multiple-choice format, requires a 'best fit' analysis by which each alternative is matched against the text and the one with the greatest overlap with the text premises selected. Thus, the more subtle the differences among alternatives, the harder the selection task. A second factor which affects difficulty is the degree to which key components of the correct alternative match key components of text premises. The more direct the matches, the easier the task.

C. Semantic Selection

Task: Determine one or more semantic dimensions of a word or phrase from information given in the word's context.

Example

Question: Which types of irregularities are not relevant to floor preparation as defined by the passage?

a. shape
b. color
c. contour

Passage: Before laying asphalt tile, all floor irregularities must be removed. Otherwise, the

tiles will tend to mold themselves to them.

D. Argumentation

Task: Recognize the following relationships between a passage and a statement:

1. The statement follows from or supports one or more passage premises.
2. The statement contradicts one or more passage premises.
3. The statement can be deduced from the passage premises.
4. The statement is a generalization of the passage premises.

This is a complex area of inference that can not be completely explicated in this paper. The general form of argumentation questions should be as shown in the example below.

Instructions: Read the following passage; then, determine for the statements which come after it (1, 2, 3) whether each follows or does not follow from what is said in the passage.

A sociologist surveyed, by means of a mail questionnaire, the attitudes of persons who managed a certain group of hotels and restaurants as to whether they would accept Chinese as guests or customers. He then arranged for a Chinese couple to visit these hotels and restaurants, and subsequently learned from the couple which establishments had actually served them. He found that of the establishments which had served the Chinese couple, over 90 per cent had previously stated they would not serve Chinese.

1. Expressed attitude toward a course of action is not necessarily a reliable indicator of behavior.
2. Surveys measuring expressed attitudes contribute nothing to the understanding of what people will do in everyday practice.
3. The majority of the managers of hotels or restaurants which served this couple during their travels had said they would refuse to accept Chinese as guests or customers.

This example tests generalizations (statement 1 & 2) and restatement (statement 3). Interpretation of argumentation statements often hinges on proper understanding of noun quantifiers (all, none, nothing, some, few, most, etc.) and verb qualifiers (almost always, not necessarily, seldom, never, etc.) The degree of difficulty of such tasks varies with the difficulty of applying these terms, plus the role played by world knowledge.

Clearly, difficulty can be increased by increasing the amount of knowledge which the reader must supply. This may be inappropriate, however, for a national survey. If reader knowledge is to be considered, it should only be in a negative sense: that is, in terms of the reader's ability to eliminate interference from world knowledge. Consider, for example, the following item, which would be introduced exactly as the one just given above.

Example: Usually I fall asleep promptly, but about twice a month I drink coffee in the evening; and whenever I do, I lie awake and toss for hours after I go to bed.

1. My problem is mostly mental; I am over-aware of the coffee when I drink it at night, anticipating that it will keep me awake, and therefore it does.
2. I don't fall asleep promptly after drinking coffee at night because the caffeine in coffee stimulates my nervous system for several hours after drinking it.
3. Whatever causes me to lie awake and toss at night is associated with my drinking coffee earlier in the evening.

Statement 2 might be selected on the basis of common (world) knowledge, but does not follow from the text premises.

Subjects might be requested to scale argumentation statements (e.g., true, probably true, insufficient data, probably false, false). In addition, interpretation of the statements (and of text premises) should hinge mostly on correct interpretation of quantifiers and qualifiers. Negatives and complex logical relations should be used to increase task complexity.

Synthesizing Operations

A. Outlining

Task analysis. Outlining involves the following tasks:

a. Logical segmentation of text into topics and subtopics
b. Ordering of subdivisions (topics and subtopics) into a hierarchy
c. Generation of suitable labels for each unit
d. Proper presentation of labels.

Of these tasks, d is purely mechanical and involves proper use of Roman numerals, letters, etc., plus indentation and titling. Tasks a and b can vary in difficulty according to the organization of the text. If all topics and subtopics are totally separate (i.e., do not interrupt each other) and if the text order

mirrors the outline order, minimal difficulty should occur. (Some ambiguity might occur, however, if the student assumes that an outline should mirror the text ordering, regardless of the logical relationships among topics and subtopics.) Task c seems relatively unimportant at first glance, but might be equal in difficulty to tasks a and b.

Difficulty levels. Texts should vary from highly familiar (i.e., common) to highly uncommon contents. The degree to which topics and subtopics are overtly marked (topic sentences, etc.) should also vary, but should in general tend toward the overt end of the spectrum.

B. Summarizing

Of the Synthesis operations, summarizing appears to be the most difficult (Anderson & Armbruster, 1984). It has been studied as a learning strategy for reading comprehension (e.g., Hare & Borchardt, 1984), but little has been written on the cognitive processes required by the task. Alvermann & Moore (1991) speculate that one component of the difficulty in summarizing derives from students' inability to identify the important propositions in a text. For the assessment proposed here, summarizing is a critical skill to consider. A variety of approaches might be adopted for item development, but each will require extensive pilot testing.

Task 1. Summarize a sequence of non-verbal instructions. These might, for example, be multiple-frame instructions for assembling a device (e.g., backyard swing) or using a machine.

Task 2. Summarize a comic strip.

Task 3. Summarize a debate between members of Congress over a particular bill.

In Task 3, difficulty level can be adjusted through the complexity of the debate topic selected and the nature of the opposing arguments.

Delphi examples

The primary challenge here is to integrate more traditional analyses of text-based operations--summarizing, outlining, making inferences, etc.--with the critical thinking categories. To the degree that this can be done successfully, literacy assessment will be distinguishable from content area assessment and from assessment of thinking skills by themselves.

The Delphi classification of core critical thinking skills and sub-skills is as follows:

1. Interpretation
   Categorization
   Decoding significance
   Clarifying meaning

2. Analysis
   Examining ideas
   Detecting arguments
   Analyzing arguments

3. Evaluation
   Assessing claims
   Assessing arguments

4. Inference
   Querying evidence
   Conjecturing alternatives
   Drawing conclusions

5. Explanation
   Stating results
   Justifying procedures
   Presenting arguments

For the analysis that follows, only the first category (Interpretation) is explored.

Interpretation: Categorization

Categorization tasks are common in scientific and bureaucratic writing where a variety of classes are defined in terms of values for a set of parameters. Exemplars are then presented and according to their features, placed in appropriate categories. For example, part of the Internal Revenue Services Circular E: Employer's Tax Guide (Pub. 15; Rev. January 1986) is shown below 12-13).

Special classes of employment and especial types of payment	Treatment under different employment taxes		
	Income tax withholding	Social security	Federal unemployment
Agricultural labor	See Circular A	See Circular A	See Circular A . . . . .
Aliens: (Aliens not qualifying as resident aliens under the statutory definition contained in IRC section 7701(b) are defined as nonresident aliens. See Publication 519, U.S. Tax Guide for Aliens, for more information.)			
a. Resident			
1. Service performed in U.S. . .	Same as U.S. citizen	Same as U.S. citizen; service as crew member of foreign vessel or aircraft exempt if any part is performed outside U.S.	Same as U.S. citizen
2. Service performed outside U.S.	Same as U.S. citizen	Taxable if (1) working for an American employer or (2) an American employer by agreement covers U.S. citizens and residents employed by its foreign affiliates.	Exempt unless on or in connection with an American vessel or aircraft and either performed under contract made in U.S., or alien is employed on such vessel or aircraft when it touches U.S. port.
b. Nonresident working in U.S.:			
1. Canadians and Mexicans entering U.S. frequently in transportation service across boundary, or in construction or operation of waterway, bridge, etc., at boundary.	Exempt under the conditions stated in the regulations.	Exempt if railroad service or if covered by totalization agreement.	Exempt if railroad service
2. Other Canadians and Mexicans entering U.S. frequently to work.	Same as U.S. citizen	Same as U.S. citizen but exempt if covered by totalization agreement.	Same as U.S. citizen
3. Workers from any foreign country or its possession lawfully admitted on a temporary basis to perform agricultural labor.	Exempt when performing agricultural labor.	Exempt	Taxable starting in 1986, same as U.S. citizen.
4. Student, scholar, trainee, teacher, etc., as nonimmigrant alien under section 101(a)(15)(F) or (J) of Immigration and Nationality Act.	Taxable unless excepted by regulations.	Exempt if service is performed for purpose specified in section 101(a)(15) (F) or (J) of Immigration and Nationality Act.	
5. All other nonresidents working in U.S.	Taxable unless excepted by regulations.	Same as U.S. citizen; service as crew member of foreign vessel or aircraft exempt if any part performed outside U.S. and employer is not "American employer." Exempt if covered by a totalization agreement.	Same as U.S. citizen

A variety of classification tasks could be generated around this table. For example, we might ask what Federal unemployment tax each of the following should pay:

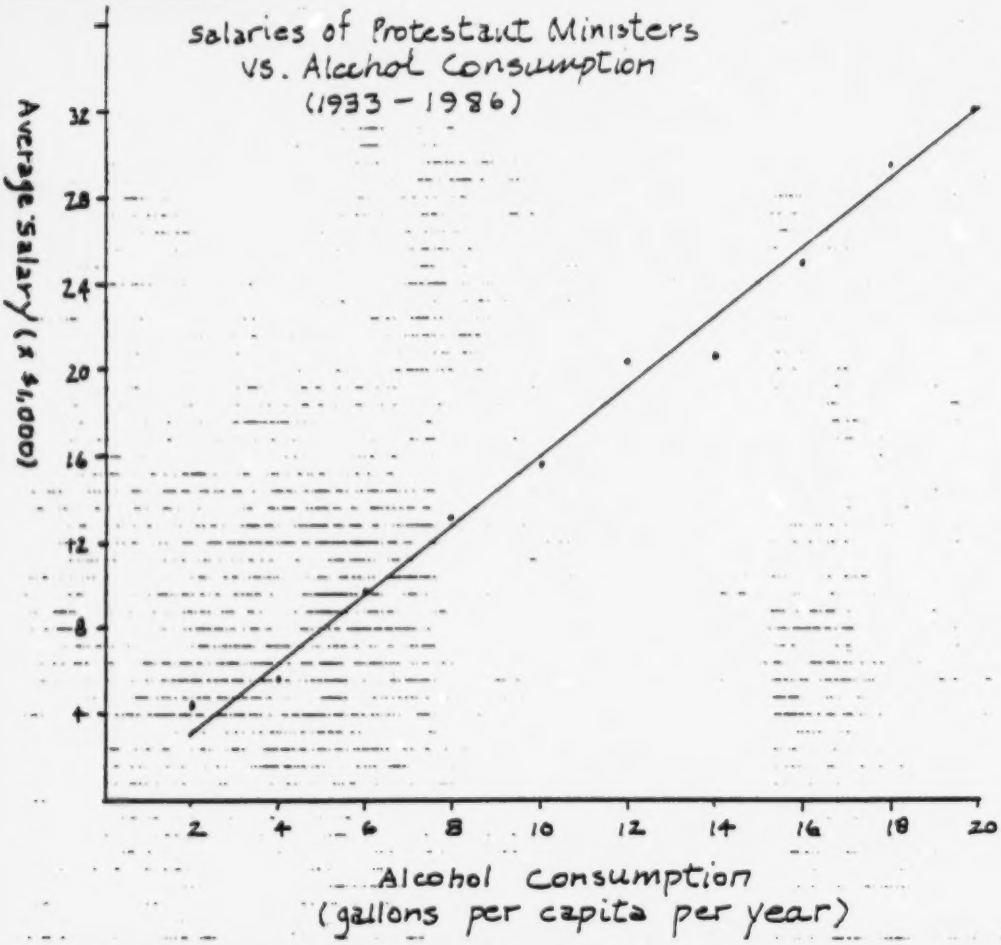
1. A nonresident Canadian who enters the U.S. several times each week to work on a barge canal.
2. A Costa Rican who enters the U.S. several times each week to work on a barge canal.
3. A nonresident Mexican who enters the U.S. several times each week while working on a railroad freight line.

Difficulty levels can be adjusted by use of synonyms and other forms of indirect statement.

Interpretation: Decoding significance

BEST COPY AVAILABLE

One type of item for assessing decoding of significance is represented by the graph below which plots the relationship between alcohol consumption in the United States and the average salaries of Protestant ministers.

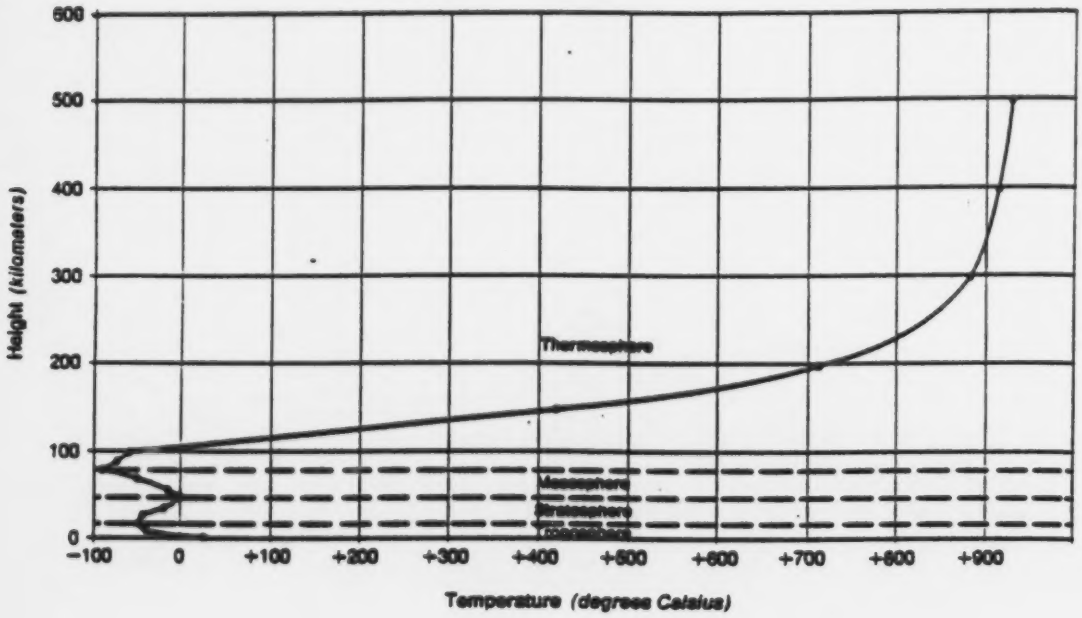


An adequate interpretation of the significance of the graph should include the following:

1. In general, as alcoholic consumption increases, so does the average salary of Protestant ministers;
2. The relationship between the two variables appears to be linear; and
3. Neither the interpretation that paying ministers higher salaries causes people to drink more alcohol, nor the reverse, that higher alcohol consumption causes higher minister salaries is a valid conclusion from the graph.

Interpretation: Clarifying meaning

A variety of different types of items could be developed for this category, using both verbal and graphical materials. For example, the graph shown below plots the temperature of the atmosphere at different altitudes above the earth's surface. One task might be to describe in words the content of the graph, where an adequate response would not require special knowledge of molecular theory or the properties of outer space.



Integration & analysis

In this section I will demonstrate problems in the assessment of critical comprehension by using a single sentence from Lincoln's Gettysburg Address. Many of the ideas reflected here draw upon interpretations by Brann (1976) and Thurow (1976) of the Address. The sentence in question is the opening sentence which reads as follows:

"Four score and seven years ago our fathers brought forth on this continent

a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal."

What I will demonstrate are different levels or depths of analysis upon which college-level assessment might be constructed. These levels are lexical/syntactic, propositional, interpretive, and critical.

A strictly lexical/syntactic analysis would focus on the specific vocabulary and the syntactic relationships between lexical items. For example, score would have to be equated with twenty years; four score and seven years would then have to be equated with 87. Our fathers would have to recognized as the subject of the sentence, brought forth as the main verb, a new nation as the object of the verb. Then, the two clauses conceived in Liberty and dedicated to the proposition that all men are created equal would have to be recognized as modifying nation. Critical to the higher levels of interpretation which follow are the correct meanings of the terms dedicated and proposition.

The second level of analysis would deal with recognition of the various propositions in this sentence, of which there are at least three. The first and main proposition is that someone did something at a particular time and place, meaning that our fathers brought forth a new nation 87 years ago and at a place that was identified by Lincoln only as on this continent. The second proposition is that the new nation was conceived in liberty and the third is that the new nation was dedicated to a particular proposition, that proposition being that all men are created equal. The phrase all men are created equal might be treated as a fourth proposition, but that's not a critical concern at this point. The isolation of propositions probably represents an initial level to begin to define college-level assessment of comprehension. From this point, one could branch to a stylistic analysis and ask such questions as why Lincoln used four score and seven years ago rather than just simply 87 years ago, but this is an issue to be covered in a totally different area of assessment and I won't pursue it further here.

The next level of analysis is what I would call an interpretive level and it requires questioning what these various propositions mean. It demands, therefore, the integration of information from the sentence with previously acquired knowledge. To begin we might ask what happened 87 years ago. To answer this we must know that Lincoln's Gettysburg Address was delivered in November of 1863 and therefore 87 years prior to that point in time was the year 1776. One could stop at this point and say, "Well, fine; Lincoln assumed that the U.S.A. was founded in 1776," but our knowledge of history leads us to associate the founding of new nations with specific events: a peace treaty, a constitutional assembly, an international charter of recognition, or the like. For the United States, this event might have been the adoption by the Continental Congress on July 2, 1776, of a

resolution put forward by Richard Henry Lee, declaring that the United Colonies "are, and of right ought to be, free and independent States." An alternative hypothesis, and a more likely one given the remainder of the sentence, is that Lincoln had in mind the Declaration of Independence, which was adopted by the Continental Congress on July 4, 1776.

At a critical level of analysis we can ask why many southerners reject Lincoln's claim that this country began with the Declaration of Independence, and not the Constitution or perhaps the Mayflower Compact. An adequate answer to this question would need to be built on Lincoln's concern with equality. Neither the Constitution nor the Mayflower Compact provide a basis for his claims of equality, but the Declaration of Independence does.

A second point of interpretation centers on the phrase our fathers. What does our fathers mean, given that the majority of the citizens of the United States in 1863 were immigrants whose fathers were not on this continent in 1776 when the Declaration of Independence was adopted. Is our fathers equivalent to our forefathers? Does it refer specifically to the signers of the Declaration of Independence or is it referent the entire group of colonials who assisted in the separation from England? An adequate answer to this question requires some knowledge of how Lincoln saw citizenship in the United States. Lincoln may have assumed, for example, that all citizens, whether they were immigrants or natives who came over on the Mayflower, inherited the full history of the country, so that our fathers is not meant in a literal sense but much more in a metaphorical sense. That is, the founding fathers were for all citizens fathers in that they created the very nation within which they then held citizenship.

The third focus of attention in this analytical exercise is on the phrase dedicated to the proposition that all men are created equal. A simple interpretation of this phrase, bringing to bear what knowledge a college graduate should have of the Declaration of Independence, should lead him or her to sense that something is wrong here. The Declaration of Independence does not say that we are dedicated to the proposition that all men are created equal; instead it says that "We hold these Truths to be self- evident, that all Men are created equal." Why then is Lincoln saying that the country was founded with a dedication to this particular proposition rather than with an assumption of this truth? To answer this we have to elevate our analysis to a critical level and look more carefully at the conflict in the Declaration of Independence between this self-evident truth as stated in the introduction and the later stated necessity of governments to derive their powers from the consent of the governed. What if the governed do not consent to the assumption of equality? Certainly the drafters of the Constitution did not, at least not for slaves and women. Lincoln's way around this conflict was to claim that this country was founded with a dedication to a not yet achieved goal of equality.

All four levels of analysis reflected here: lexical/syntactic, propositional, interpretive, and critical, are potential bases for literacy assessment at a post-secondary level. The first two-- lexical/syntactic and propositional--do not require prior knowledge of American history for adequate responses, but the remaining two do. Could the analyses outlined here be reasonably required of college graduates? Certainly not all of the interpretive and critical analyses, but perhaps some of them, provided that adequate supporting materials were provided.

Note that even for history majors it would be unreasonable to expect the availability from memory of the information required here on American history. Thus, if interpretive and critical analyses are desired, they will need to be built upon tasks which allow access to background documents. The total task would then require good information retrieval skills, rapid skimming ability, and a high rate of silent reading, as well as the interpretive and critical skills around which the present task was conceived.

Arguments for/against Definitions of literacy

Although the functional definition of literacy such as those used by the National Assessment of Educational Progress and Statistics Canada seem totally appropriate for national surveys of literacy where one's focus is primarily on lower levels of functioning, it is not equally obvious that functional definitions are appropriate for an assessment of higher education. Current directions in testing of cognitive abilities (e.g., Baker, Freeman, and Clayton, 1991) focus on "important and teachable processes" (p. 135), which implies that assessment should provide appropriate clues for instruction. Although any cognitive model of literacy competence will change over time, it is probably more reasonable to acknowledge this possibility and nevertheless attempt to define levels of expertise based upon human information processing rather than to anchor literacy assessment on current demands of societal functioning.

One reason for eschewing the latter is that we do not know at any point in time whether the "demands of society" are reasonable and appropriate demands for human capabilities. Take for example the literacy demands in manufacturing. For many years American manufacturing was uniformly based upon a specialization model that minimized skill demands of the individual workers. Each worker was trained in an narrow range of functioning and then was employed to apply repeatedly some subset of available skills. In contrast, a number of industries have moved towards a more group based approached to manufacturing where clusters of workers take full responsibility for the total set of operations required for the production of some object. This latter approach requires a higher level of skills than the former, not only for the specialized manufacturing operations, but also for communication

and inter-personal relations. Under the first organizing principle for manufacturing, a set of literacy skills would be derived, but these would be far too low for what would be required for the second approach.

A further reason for rejecting the functional approach to defining higher levels of literacy is that it is in opposition to the traditional values of liberal education in America. Within this tradition, the goal of education is to prepare critical, reflective citizens who can make judgments on their own; in other words, to prepare the inner-directed person. To define literacy in terms of the print demands of any given level of contemporary society is to make the individual subservient to the whims of the masses. Liberal education aspires to lift limits on individual accomplishment rather than to impose them.

Therefore to be compatible with current directions in cognitive assessment, it is recommended that literacy assessment be based upon criterion referenced measurement and that it focus upon those skills that from the best knowledge available we assume are required for expert behavior in literacy.

Skill Types

Arguments for or against inclusion of particular skill types are difficult to justify without knowledge of other proposals for critical thinking assessment. Of the material presented in the previous section, the ideas derived from the analysis of Lincoln's Gettysburg Address seem most central to the entire enterprise. That is, the ability to evaluate and to analyze critically a passage are important outcomes of a college education. Lower level skills (e.g., predicting outcomes, locating information) may be useful diagnostically, but a higher level literacy assessment would be incomplete without evaluation and critical analysis items.

The forthcoming National Adult Literacy Survey (NALS) may provide information on the types of lower level skills which college graduates still have not mastered. This will require, however, a careful task analysis of the NALS assessment exercises. The S.A.T. may also provide such information, although the students who take this test are not yet in college.

A literacy assessment, if it taps critical analysis ability, will require both reading and writing. Furthermore, I can see no reason to exclude charts, graphs, and other visual representations. The critical issue is to find or generate representative texts for more extended tasks -- summarizing, outlining, etc.-- and these might contain a full range of data representations. What I would suggest excluding is creative writing, primarily because the research base for this ability is severely limited and the number of students who receive instruction in it while in college is small.

Separate vs. embedded assessment

After wandering through the presentation on assessment options, I am convinced that lower level reading skills can be evaluated independently of specific curricular areas but that higher level literacy skills probably cannot. At the highest levels of literacy assessment are those skills that require integration of text-derived information with information obtained previously or from other texts in the same task. Although artificial situations might be created, drawing upon "neutral" content, I doubt that meaningful items can be constructed by this means. Therefore, to eliminate some of the confounding of content area competence with literacy ability, I suggest that some of the higher level tasks provide content area background materials along with the test passages, and that these be located primarily in each student's major area.

References

Adult Performance Level Project. (1977). *Final report: The Adult Peformance Level Study.* Washington, DC: U.S. Office of Education.

Alvarez, Leo Paul S. de (Ed.) (1976). *Abraham Lincoln: The Gettysburg Address and American Constitutionalism.* Irving, Texas: University of Dallas Press.

Alvermann, D.E., & Moore, D.W. (1991). Secondary school reading. In R. Barr et al. (Eds.), *Handbook of reading research* (Vol. 2). NY: Longman.

Anderson, T.H., & Armbruster, B.B. (1984). Studying. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal, (Eds.), *Handbook of reading research* . White Plains, NY: Longman.

Baker, E.L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition.* Englewood Cliffs, NJ: Prentice Hall.

Barr, R., Kamil, M.L., Mosenthal, R., & Pearson, P.D. (Eds.) (1990). *Handbook of reading research* (Vol. 2). NY: Longman.

Birkmire, D.P. (1982). *Effect of the interaction of text structure, background knowledge, and purpose on attention to text.* Unpublished doctoral dissertation, University of Delaware.

Bond, G.L., & Bond, E. (1941). *Development of reading in high school.* NY: Macmillan.

Brann, E. (1976). A reading of the Gettysburg Address. In L.P.S. de Alvarez (Ed.), *Abraham Lincoln: The Gettysburg Address and American Constitutionalism.* Irving, Texas: University of Dallas Press.

Brown, A.L. (1987). Metacognition, executive control, self-regulated and other more mysterious mechanisms. In F. Weinert & R. Kluwe (Eds.), *Metacognition, motivation, and understanding.* Hillsdale, NJ: Erlbaum.

Brunner, E. de S., Wilder, D.S., Kirschner, C., & Newberry, J.S. (1959). *An overview of adult education research.* Chicago: Adult Education Association of the U.S.A.

Buswell, G.T. (1937). How adults read. *Supplementary Educational Monographs, 45.* Chicago: University of Chicago Press.

Chall, J.S. (1983). *Stages of reading development.* NY: McGraw Hill.

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser.* Hillsdale, NJ: Erlbaum.

Dixon, W.R. (1951). Studies of the eye movements in reading of university professors and graduate students. In W.C. Morse, F.A. Ballantine, & W.R. Dixon (Eds.), *Studies in the psychology of reading. Monographs in Education, 4.* Ann Arbor, MI: University of Michigan Press.

Facione, P.A. (1990). Executive summary of "Critical thinking: A statement of consensus for purposes of educational assessment and instruction" Millbrae, CA: California Academic Press.

Final Report: (1977). *The Adult Performance Level Study*. Austin, TX: Adult Performance Level Project.

Fischer, J.K., Haney, W., & David, L. (1980). *APL revisited: Its uses and adaptation in states*. Washington, DC: Office of Educational Research & Improvement. U.S. Department of Education.

Glaser, R., & Klaus, D.J. (1962). Proficiency measurement: Assessing human performance. In R. M. Gagne (Ed.), *Psychological principles in system development*. NY: Holt, Rinehart & Winston.

Glaser, R. (1991). Expertise and assessment. In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition*. Englewood Cliffs, NJ: Prentice Hall.

Goldstein, H. (1940). *Reading and listening comprehension at various controlled rates*. (Contributions to Education, No. 821). NY: Bureau of Publications, Teachers College, Columbia University.

Goody, J., & Watt, I. (1968). The consequences of literacy. In J. Goody (Ed.), *Literacy in traditional societies*. Cambridge, England: Cambridge University Press.

Gray, W.S. (1937). The nature and organization of basic instruction in reading. In W.S. Gray (Ed.), *Thirty-sixth yearbook of the National Society for the Study of Education, Part I*. Bloomington, IL: Public School Publishing.

Gray, W. (1956). *The teaching of reading and writing*. Chicago: Scott, Foresman and Co.

Griffith, W., & Cervero, R. (1977). The Adult Performance Level Program: A serious and deliberate examination. *Adult Education, 27*(4), 209-224.

Guthrie, J.T. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly, 23*, 178-199.

Hare, V.C., & Borchardt, K.M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly, 20*, 62-78.

Harris, L., & Associates. (1970). *Survival literacy study*. Washington: The National Reading Council.

Harris, L., & Associates (1971). *The 1971 national reading difficulty index: A study of functional reading ability in the United States, for the National Reading Center*. Washington, DC: The National Reading Center.

Judd, C.H., & Buswell, G.T. (1922). Silent reading: A study of the various types. *Supplementary Educational Monographs, 23*. Chicago: University of Chicago Press.

Kirsch, I.S., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults--final report.* Report No. 16-PL-01. Princeton, NJ: National Assessment of Educational Progress.

Kirsch, I.S., & Mosenthal, P.B. (1988). *Understanding document literacy: Variables underlying the performance of young adults* (Research Report No. RR-88-62). Princeton, NJ: Educational Testing Service.

Kirsch, I.S., & Mosenthal, P.B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*(1), 5-30.

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. *Reading Research Quarterly, 17,* 400-420.

Nafziger, D.H., Thompson, R.B., Hiscox, M.D., & Owen, T.R. (1976). *Tests of functional adult literacy: An evaluation of currently available instruments.* Portland, OR: Northwest Regional Educational Laboratory.

National Assessment of Educational Progress. (1976). *Adult work skills and knowledge: Selected results from the first national assessment of career and occupational development.* Denver: National Assessment of Educational Progress.

Neice, David, & Adsett, Margaret (1991). *Direct versus proxy measures of adult functional literacy: A preliminary re-examination and modeling.* Working paper, Social Trends Analysis Directorate, Department of the Secretary of the State of Canada, Ottawa.

Neijs, K. (1961). *Literacy primers: Construction, evaluation and use.* Paris: UNESCO.

Northcutt, Norvell, Kelso, C., & Barron, W.E. (1975). *Adult functional competency in Texas.* Austin, TX: University of Texas Press.

Paris, S.G., Lipson, M., & Wixson, K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8,* 293-316.

Pearson, P.D. (Ed.) (1984). *Handbook of reading research.* NY: Longman.

Rogoff, B., & Lave, J. (Eds.) (1984). *Everyday cognition: Its development in social context.* Cambridge, MA: Harvard University Press.

Scribner, S., & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.

Sheehan, K., & Mislevy, R.J. (1990). Integrating cognitive and psychometric models to measure document literacy. *Journal of Educational Measurement, 27*(3), 255-272.

Stedman, L.S., & Kaestle, C.F. (1986). *An investigation of crude literacy, reading performance, and functional literacy in the United States.* Madison, Wisconsin: Wisconsin Center for Education Research, University of Wisconsin.

Thomas, A.M. (1983). *Adult illiteracy in Canada--A challenge.* Occasional Paper #42, Canadian Commission for UNESCO, Ottawa.

Thorndike, E.L. (1917). The understanding of sentences: A study of errors in reading. *Elementary School Journal, 17*, 98-114.

Thurow, G.E. (1976). The Gettysburg Address and The Declaration of Independence. In L.P.S. de Alvarez (Ed.), *Abraham Lincoln: The Gettysburg Address and American Constitutionalism*. Irving, Texas: University of Dallas Press.

UNESCO. (1957). *World illiteracy at mid-century*. Paris: UNESCO.

Venezky, R.L., Kaestle, C.F., & Sum, A.M. (1987). *The subtle danger*. Princeton, NJ: Educational Testing Service

Venezky, R.L., Wagner, D., & Ciliberti, B. (1990). *Toward defining literacy*. Newark, DE: IRA.

Venezky, R.L. (1991). Catching up and filling in: Literacy learning after high school. In J. Flood, J. Jensen, D. Lapp, & J.E. Squire (Eds.), *Handbook of research on teaching the English language arts*. New York: Macmillan.

Review of Papers for NCES Workshop on Goal Five: Assessing Thinking and Communication in College Graduates

Robert Calfee
Stanford University
November 21, 1991

INTRODUCTION AND OVERVIEW

This memo reports on three papers prepared for the November workshop: Ewell and Jones Actions matter; Lenth Context and policy requisites, and Venezky Literacy. The memo begins with background on my approach to the review, followed by a summary and critique, and ends with a section on other issues and recommendations that occurred to me during the review process.

The three papers take different approaches and contain different substance. Given the criteria promulgated by NCES, I have focused on those elements with most direct relevance to the specifics of an assessment program. In my recommendations, I urge the workshop to give greater emphasis on writing as a primary indicator, to weigh the use of portfolio approaches as an assessment tool, and to rely on informed teacher judgment for evaluation and reporting of outcomes.

BACKGROUND

Two segments of the September 16 NCES project memo provide the background for my review. In the covering note, the goals of review are listed as (a) establishing the reliability and validity [sic] of the position papers; (b) identifying additional issues; and (c) framing the workshop agenda. The attachment on "Evaluation Criteria" includes one general point -- conceptual soundness -- as of primary importance. Five detailed criteria are also listed: (a) identifiable outcomes; (b) validity; (c) value added, (d) methods for accurate and informative assessment; and (e) practicality.

Taking these criteria as a whole, it seemed to me most important to speak to the pragmatics of post-secondary assessment as related to Goal Five: Ability of college graduates to think critically, communicate effectively, and solve problems in the workplace and in the practice of citizenship. A second theme, less clear in my reading, had to do with the "validity" of the position papers in framing the issues. I am not sure that the conditions of the task are adequate to support this rather daunting challenge. An adequate answer to Goal Five might require followup of graduates in the workplace and in citizenship activities for several years after

graduation -- I doubt that the political drive behind Goal Five is sufficient to support genuinely "valid" proposals of this sort. At the other extreme, it seems unlikely that the workshop will focus on development of multiple-choice tests of "basic literacy skills" to be mandated upon all college graduates.

Somewhere in the mid-range of these possibilities are techniques that can generate useful information not easily subject to manipulation and misinterpretation. My perspective in this review has been to explore such possibilities in the three papers, and to add a few thoughts from my own experience with assessment -- which ranges from kindergarten through the evaluation of teacher candidates, from research on standardized tests to exploration of informal assessment methods in classroom settings.

Venezky, Assessing literacy, addresses most directly of the three papers the question of what is to be assessed and how to assess it. Venezky and I have managed "parallel play" for a quarter century, so it is not surprising that I would find many connections in his paper -- and much to argue with.

"What is "literacy?" What aspects of literacy warrant assessment in the present context? What is the appropriate context for assessment? These three questions frame Venezky's paper. Literacy has to do with reading and writing, according to Venezky. The aspects worth assessment include the identification, analysis, and synthesis of text-based information. Assessment should be embedded in the content-area contexts that comprise reality for college students. I believe this synopsis captures the spirit of Venezky's argument; if I am off the mark, he will enlighten us with his clarifications.

Research on the role of printed documents in human thinking has a long history; indeed, contemporary experimental psychology built the exploration of this phenomenon, beginning with Cattell's studies of word perception, moving through Thorndike's explorations of comprehension as reasoning, past Miller's investigations of sentence grammar, and on to today's emphasis on expertise in text analysis. The text is the

stimulus, the reader is the subject, and reaction to the text is the response.

A "reading" of Venezky's survey of this history requires substantial background knowledge, but he provides useful linkages to important sources. He switches smoothly from the cognitive analysis of reading to the investigation of adult literacy. To be sure, his audience may not realize that they have passed through a time-space warp. Cognitive studies entail careful control over the situation, precise measurement of outcomes, multiple replications of the phenomenon, and a rich theoretical background guiding the research. Adult literacy, in contrast, is a mishmash of situations (mostly complex), outcomes (mostly messy or simplistic), replications (relatively few), and theory (little to speak of). Cognitive investigations require educated subjects (college sophomores, by and large) to perform laboratory tasks (mostly silly). Adult literacy research focuses on a range of backgrounds (but often people who are down and out) to do simple tasks (but not too silly) in "real" situations (wherever the researcher can locate participants).

On page 12, Venezky moves from his research review back to the main theme -- how to assess adult literacy, defined as the ability to interpret a text in some purposeful context. Unlike the other papers, by the way, Venezky gives little attention to the college graduate. His discussion stretches across the range from the unemployed to the sophomore subject, but more the former than the latter.

The material on page 12 ff left me in a puzzlement. On the one hand, in my earlier life as an experimental psychologist, I appreciated the detailed decomposition of the process of textual analysis. On the other hand, my reaction was, "Who cares?" The stimulus is rather Ebbinghausian, minute and remote from the vision of Goal Five. Recognition is required more than production, and multiple-choice testing is the antithesis of both thinking and communication, in my opinion.

By page 19, the tasks became more interesting, more for substance than for style. To be sure, the form of the argument makes it difficult to see that Venezky has shifted ground. And terms like <u>outline</u> evoke routines that, by Venezky's own judgment, may be "purely mechanical."

Venezky concludes the paper with a brief segment on <u>Arguments for/against</u> (pp. 28ff). He incorporates several meaty points in this afterword.

> "Assessment should provide appropriate clues for instruction" All three papers touch on this issue, by the way, and I recommend that the workshop give it careful consideration.

"Clusters of workers [now] take responsibility [for a job]." This point is presented not in the context of the "graduate," but of life after graduation. Venezky also addresses here the aim of college education, which is the individual's adaptation to work and citizenship.

"Literacy assessment [should] be based on criterion-referenced measurement" -- I haven't the foggiest idea what Venezky means by this remark, but the most obvious link to my background knowledge flies in the face of everything that he argues beforehand.

"Literacy assessment, if it taps critical analysis ability, will require both reading and writing." Also charts, graphs, visual representations. Why not speaking and listening? Ewell-Jones include writing in their lists, but only in their mention and Venezky's does this competence appear.

"Embedded assessment" -- the prose in this paragraph does not quite capture the spirit of what needs to be said here. College graduates are pretty good crap detectors. They can tell whether something matters. They are also serious about their "business." Kindergartners may be willing to show that they can "read" at the end of the school year. College students have specialized. They are becoming experts at something. The mathematician looks puzzled when reading Heart of darkness; the litterateur cannot comprehend Mathematical analysis. Only education students are expected to know everything. I wish that Venezky had said more about the importance of embedded assessment, which strikes me as an essential issue for college situations.

In summary, Venezky's paper spans a wide range of issues: the core questions of what is literacy, what skills support this accomplishment, and how might these skills be assessed. His focus seems unnecessarily narrow to encompass thinking and communication, but the past record has been equally narrow. His article bares an inconsistency in the current agenda -- broad scope but limited resources.

(Editors Note: Please see R. Calfee's review of Ewell/Jones' position paper for final comments)

Review of Richard L. Venezky's "Assessing Higher Order Thinking and Communication Skills: Literacy"

The paper proposes higher-order literacy assessment as a way of measuring both communication and critical-thinking skills. However, there is little analysis of the set of sub-skills contained in the first. If we assume that they are, minimally, writing, speaking, listening, and reading, literacy assessment addresses only the last. At the end of the paper, Dr. Venezky suggests an assessment that would include writing, but since no other mention is made of it or speaking, it is hard to determine how this approach constitutes an adequate measurement of communication skills.

The paper does contain a curious suggestion that reading ability might be determined if students were to both read and listen to "equivalent" passages and their understanding of the information assessed. The difference between comprehension of the two types of input would then be an "indicator of teachability," by which he seems to mean reading competency. Besides the problems with this approach mentioned in the paper (i.e. text difficulty and intelligence complicate the comparison), aural processing differs in kind from literate processing: material needs to be organized differently in each, and the skills necessary to process each kind of input are different. So comparisons are not only likely to be complicated; they also risk falling into a fundamental category mistake.

His analysis of text-based operations suggests that a link might be better made, if "with some squeezing," between critical thinking skills and traditional analyses of text-based operations. Hence those analyzes might well provide some guidance in the construction of critical-thinking assessments. Such assessments, after all, will almost inevitably involve some reading, particularly if they are not to rely on previous knowledge. It is hard to understand in what "assessment of thinking skills by themselves" could consist, in fact.

But the existing reading comprehension analyzes described in the paper would not be sufficient to the task of analyzing critical-thinking skills. First, they would both have to be both squeezed and expanded to cover the full range of skills distinguished in the Delphi classification of core critical thinking skills and sub-skills. Second, performance measures would have to be added for "authentic" assessment.

Dr. Venezky's paper does not deal with the acquisition of reading skills or how college contributes to that acquisition, although his rejection of the "functional" approach to defining higher levels of literacy in favor of the cognitive approach implies that higher education should, in some way unspecified, teach certain thought processes. And he argues, finally, that since higher-level literacy assessment probably cannot "be

evaluated independently of specific curricular areas," background materials should be supplied along with the test passages and "that these be located in each student's major area." This, of course, lands him in the middle of the controversy over whether critical thinking can be independent of specific disciplinary mastery. The solution of the framers of New Jersey's GIS examination -- to assume that certain discipline-linked intellectual skills will characterize all generally educated persons and hence to supply background materials but not to link the test to the student's major -- seems preferable.

In short, Dr. Venezky's paper does suggest that traditional literacy assessment may have something to offer in the assessment of critical thinking skills and how to differentiate tasks to measure various of the sub-skills that make up critical thinking. But he has little to say about how these skills are learned or not learned. And a significant and expensive amount of research has yet to be done to develop performance evaluations of the full range of critical-thinking skills.

Margaret A. Miller

PACIFIC GRADUATE SCHOOL OF PSYCHOLOGY

935 EAST MEADOW DRIVE, PALO ALTO, CALIFORNIA, 94303

(415) 494-7477

Evaluation Institute

Michael Scriven, Director

COMMENTS

3. VENEZKY ON LITERACY ASSESSMENT

The Venezky paper was very interesting in particular because of its rich supply of interesting practical examples, although not all of them are relevant to the task we face here. Three brief comments. First, I did not find it helpful to make the 'human expertise' vs. 'social demands' distinction that he thinks distinguishes CT from ordinary literacy. It seems clear enough that in both cases, and with respect to communcation and problem-solving, it is external world requirements for which we are trying to educate students; expertise is just a measure of the extent to which they master these requirements. I have set out the alternative view in some detail in "Critical Thinking & the Concept of Literacy", *Informal Logic*, Spring, 1989, based on my experience in creating and enforcing a literacy test on teacher-trainees in a school of education. The results were of some interest to our general task here. Over 40% failed the first test, but over 95% of them passed once they had a chance to see the results and work on improvement for a few months; for the most part, those who failed had simply never been pulled up for deficiencies in their school or college classes, but had no great difficulty in meeting the standards once these

were identified. (Nor did they disagree with the standards, although they were given the opportunity to do this anonymously.)

I find this consistent with results in the CT area, which is, to my way of thinking, simply an extension of literacy in the narrow sense. (I have outlined a full 'extended literacy' curriculum in "Functional English", ***English In Australia***, 1984, pp. 33–43, 67.) Of course, communication is even more clearly part of any *extended* sense of literacy, and substantial skills in problem-solving—although not all of the subject-matter linked ones-are clearly a crucial part of CT. Which brings me to my second comment: it seemed surprising that Venezky did not more extensively address speaking and listening competencies rather than the fine points of testing writing competencies. We would surely be inclined to regard it as part of communicative literacy to be able to take reasonably accurate notes from a lecture or talk, give a reasonable verbal summary from them, and respond verbally to questions about the content of the talk. Testing that I have done in adult education contexts makes clear that these skills, which are part of what we think 'must' be learnt in the process of attaining high skill graduation, are extremely variable. For example, some people are selectively deaf to points of view with which they disagree; arguments supporing those points of view, given in a talk they hear, simply do not show up in their notes on that talk or show up in an absurd form. I suspect that we need to do some analysis of ordinary (written) literacy assessment tests to see if this is not a general phenomenon, and I will begin that work on a small scale in June.

The third comment is simply that Venezky also seems to short-change the possibility of non-multiple choice but nevertheless 'objectively' scorable tests of reading literacy. I've provided some discussion of these in an appendix.